Published in final edited form as:

*Dev Cell*. 2018 December 17; 47(6): 773–784.e6. doi:10.1016/j.devcel.2018.11.019.

# A developmental program truncates long transcripts to temporally regulate cell signaling

**Jeremy E. Sandler**[1], **Jihyun Irizarry**[1], **Vincent Stepanik**[1], **Leslie Dunipace**[1], **Henry Amrhein**[1], and **Angelike Stathopoulos**[1,*]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125 USA.

## SUMMARY

Rapid mitotic divisions and a fixed transcription rate limit the maximal length of transcripts in early *Drosophila* embryos. Previous studies suggested that transcription of long genes is initiated but aborted, as early nuclear divisions have short interphases. Here, we identify long genes that are expressed during short nuclear cycles as truncated transcripts. The RNA binding protein Sex-lethal physically associates with transcripts for these genes and is required to support early termination to specify shorter transcript isoforms in early embryos of both sexes. In addition, one truncated transcript for the gene *short-gastrulation* encodes a product in embryos that functionally relates to a previously characterized dominant-negative form, which maintains TGF-β signaling in the off-state. In summary, our results reveal a developmental program of short transcripts functioning to help temporally regulate *Drosophila* embryonic development, keeping cell signaling at early stages to a minimum in order to support its proper initiation at cellularization.

## eTOC

- Long transcripts are truncated during short nuclear cycles in *Drosophila* embryos
- The RNA-binding protein Sex-lethal binds to transcripts and controls their truncation
- Short transcript products are functional in signaling pathways, affecting initiation
- Global 3’ RNA-seq identifies additional truncated transcripts suggesting a program

*Correspondence and Lead Contact: angelike@caltech.edu.

AUTHOR CONTRIBUTIONS

J.E.S. and A.S. designed the experimental approach. J.E.S. performed RACE, in situ hybridizations, RNAi, qPCR, RNA-seq, and transgenic embryo experiments. J.I. established CRISPR fly stocks and performed antibody stainings, V.S. performed immunoprecipitations and Western blots, and L.D. generated transgenic embryos and performed in situ hybridizations. J.E.S., L.D., and A.S. analyzed data and H.A. performed computational analysis on RNA-seq data. J.E.S. and A.S. wrote the manuscript with contributions from J.I., V.S., L.D., and H.A.



DECLARATION OF INTERESTS

The authors declare no competing interests.

## INTRODUCTION

Early embryonic development of the fruit fly *Drosophila melanogaster* is characterized by 14 rapid and syncytial mitotic nuclear cycles (NCs) as the fertilized egg divides into ~6000 nuclei before cell membranes form and gastrulation occurs (Foe and Alberts, 1983). These NCs occur within three hours of egg laying and vary in length from ~10 minutes to about an hour, gradually lengthening as the embryo nears gastrulation (Pritchard and Schubiger, 1996; Tadros and Lipshitz, 2009). This rapid pace of nuclear divisions leads to a dynamic transcriptional environment, where patterns and levels of gene expression change between and within NCs (Reeves et al., 2012; Sandler and Stathopoulos, 2016a). Transcription is aborted during mitosis between NCs, and nascent transcripts are degraded, with transcription restarting at interphase of the following NC (Shermoen and O'Farrell, 1991).

As the rate of transcription in *Drosophila* has been measured at ~1.1–1.5kb per minute of interphase (Ardehali and Lis, 2009; Garcia et al., 2013), transcription of zygotic genes during syncytial NCs is likely time constrained. In support of this view, early zygotic genes have an average length of 2.2kb, while the overall average length of coding genes in *Drosophila* is 6.1kb (Artieri and Fraser, 2014; Hoskins et al., 2011), suggesting a bias towards short genes during this time period. It was previously thought that long genes, those over 20kb, are either not transcribed before the longer and final syncytial NC14 or are aborted mid-transcript, and no protein products were present (O'Farrell, 1992; Rothe et al., 1992).

Activation of the zygotic genome and the maternal to zygotic transition (MZT) takes place during the syncytial nuclear period and cellularized blastoderm period before gastrulation, concurrent with time constraints on transcript length (Tadros and Lipshitz, 2009). This is also when the dorsal-ventral and anterior-posterior axes that pattern the embryo, and eventually the adult fly, are established by zygotically transcribed genes relying on a few key maternal signals. Lastly, components of virtually all signaling pathways are zygotically transcribed during this time (Lott et al., 2011; Sandler and Stathopoulos, 2016b) and these signaling pathways, such as TGF-β, JAK/STAT, Notch, FGF, and EGFR, are active and essential during embryonic development (rev. in Stathopoulos and Levine, 2004). For all these reasons, it is essential that the necessary genes for these processes be transcribed at the correct time in development, yet the observations of the exclusion of long genes remain, along with the questions about consequences for development in the absence of these transcripts.

Recently, studies have produced evidence that some long genes are transcribed during early NCs (Ali-Murthy et al., 2013; Lott et al., 2011; Sandler and Stathopoulos, 2016b). To explore these observations that seemingly contradict previous research, we examined transcription of long genes during short syncytial NCs, specifically NC13, with an interphase of 15 minutes, and compared the transcription of these same genes during the longer interphase associated with NC14, which is over 45 minutes.

## RESULTS AND DISCUSSION

### Long transcripts are truncated during short nuclear cycles

Using an available RNA-seq dataset of *Drosophila* early embryonic development, we selected four genes over 20kb with evidence of transcription during NC13: *short gastrulation (sog), scabrous* (*sca*), *Protein kinase cAMP-dependent catalytic subunit 3 (Pka-C3)*, and *Netrin-A (NetA)* (Figure 1A) (Lott et al., 2011). 5' and 3' rapid amplification of cDNA ends (RACE) was performed on RNA from embryos aged 1–3 hours, which includes NCs 13 and 14 (Figure 1B), to search for alternative transcript isoforms. Only the previously defined 5' transcription start sites were recovered (Graveley et al., 2011) suggesting that alternative start sites are not used for these genes, whereas 3' RACE products identified truncations in these four transcripts (Figure 1A). The short forms aligned to annotated transcripts at the beginning of the full-length genes, but ended with an alternative exon, including new coding sequence and a 3' UTR in what is usually an intron (Figure 1A, red transcripts; and Figure S1). The RACE products were spliced and polyadenylated, with no poly-A in the genome at the locus of alignment, suggesting they are mature transcripts.

To distinguish between full-length transcripts and short forms, we designed fluorescence in situ hybridization (FISH) riboprobes to the 5' and 3' ends of *sog, sca, Pka-C3*, and *NetA* with 3' exonic probes downstream of mapped short RACE sequences and therefore recognizing full-length forms only (Figure 1A). In all cases, there was no observable nascent signal from the 3' exonic probes during NC13 while signal from the 5' exonic probes was present, indicating that transcription did not reach the 3' ends of genes assayed (Figure 1C,D,G,H). In contrast, full-length transcripts were present in NC14 when interphase is longer, as indicated by equivalent levels of expression detected by both 5' and 3' probes (Figure 1E,F,I,J). The signals were quantified relative to ubiquitous histone staining and compared for NC13 and NC14, showing that at NC13 the signals associated with 5' versus 3' ends were significantly different while roughly equivalent at NC14 (Figure 1K–N).

### The RNA-binding protein Sex-lethal controls transcript truncation

Since the short transcripts include intron-derived coding sequence (Fig. S2A–D, blue sequences), we reasoned it is likely that transcriptional regulation is a cause of truncation at NC13 as opposed to post-transcriptional cleavage of full-length, mature mRNAs, which after splicing would lack intron-derived coding sequence. The sequence within 1 kB downstream of the new exons was examined for all four transcripts found to be truncated. While there were binding sites for 20 temporally relevant RNA Binding Proteins (RBPs) in all four genes, we found that the sites for Sex-lethal (Sxl) (Figure 2A; Ray et al., 2013) were the only ones statistically enriched, with $p<0.001$ calculated using the Anaysis of Motif Enrichment (AME) software package (see STAR Methods; Data S1) (McLeay and Bailey, 2010).

*Sxl* is a well characterized sex determination gene in *Drosophila* involved in splicing (Moschall et al., 2017; Salz and Erickson, 2010). Zygotic expression of functional Sxl protein only occurs in female embryos, while males express a non-functional form (Bell et al., 1991; Bopp et al., 1991). However, short transcripts of long genes (e.g. *sog*) were

observed in all embryos examined at NC13 (not only females) demonstrating that the RBP fulfilling this role is not sex-specific. Notably, *Sxl* is also maternally expressed with transcripts deposited into eggs and early embryos; while based on activation of the *Sxl* associated Pe zygotic promoter and in situ hybridization using riboprobes designed to the 5' end of the gene (i.e. Ex1), female-specific, zygotic transcription is thought to occur at NC11 (Cline, 1993; Keyes et al., 1992). It remains unclear, however, whether full-length transcripts of the long (>23kb) *Sxl* gene can be completely transcribed. Moreover, when we examined RNA-seq data from a fine time course of early *Drosophila* development, we found support for the view that zygotic *Sxl* transcripts are not upregulated in females until mid NC14 (Lott et al., 2011).

Since we observed short transcript production in embryos of both sexes, we investigated whether at NC13 maternal Sxl could support this role. Maternal and female-specific zygotic mRNA transcripts should support the production of proteins with shared sequences and thus be recognized by the same antibody. However, although we were able to detect female-specific Sxl protein by Western blot at NC14 and show specificity for the antibody via maternal RNAi knockdown that was also able to downregulate zygotic levels (Figure 2H and S3M,N), we were unable to unambiguously visualize Sxl in unfertilized and early stage embryos, as bands of similar (but not identical) size to female-specific Sxl identified by Western blot in these embryos and early stages were unaffected by the equivalent RNAi conditions, suggesting these are background bands that possibly masking true maternal Sxl (Figure S2M,N). As assays of maternal Sxl by Western proved inconclusive, immunostaining to examine the protein in individual embryos at NC13 did reveal Sxl present in both sexes, using an intronic probe to *sog* (on the X chromosome) to determine the sex of the embryos (Figures 2B–F). In both male and female embryos, we observe presence of Sxl protein at NC13 (Figures 2B',C' and S3O,P) and in earlier NCs as well (e.g. Figure S2K,L). Sxl levels are reduced by heat-shock induced, maternal RNAi, initiated during oogenesis but which perdures into the early embryo (Figures 2J–M; see Methods). The immunostaining of individual embryos is sensitive enough to detect low levels of Sxl and to identify that a ~40% reduction occurs upon Sxl RNAi (Figure 2G). Furthermore, this fine time resolution analysis of Sxl protein levels demonstrates that female-specific, zygotic Sxl protein is not produced until late NC14 and suggests that earlier signal detected by immunostaining relates to a low-abundance maternal isoform (Figures 2D',E' and S3Q–T). To support this, maternal Sxl at NC13 in both sexes (Figures 2B',C') is much lower level than zygotic Sxl in females at late stage NC14 (Figure 2E',S3O,P,T) but comparable to (or even higher than) levels retained at this stage in males (Figure 2D',S3P,Q,S) or in female *daughterless* (*da*) mutant embryos (Figure 2F'), where all *Sxl* zygotic transcription is eliminated (Cronmiller and Cline, 1987).

To characterize effects of maternal *Sxl* RNAi on transcript truncation, which showed sufficient knock-down of Sxl (Figures 2G, J–M, and S3N), riboprobes targeting transcription of intronic sequences 3' of the initially defined truncation sites of the four long genes, just 3' of the cluster of predicted Sxl binding sites (e.g. Intron 3 probe, Figure 2I), were used to test the hypothesis that Sxl binding to the transcript could act to influence termination. In embryos subject to maternal RNAi against *Sxl*, intronic FISH signal past the truncation point was observed during NC13 for all four long genes assayed, indicating that transcriptional

read-through past the truncation point occurs in both sexes (Figures 2N,O and S3A,C,E, compare with Figure 2P and S3B,D,F,G,J). Furthermore, *da* mutants expressing only maternal Sxl were not able to support transcriptional read-through past the truncation point (Fig. 2Q). Collectively, these results support the view that maternal, not zygotic, Sxl is responsible for transcriptional truncation in early-stage embryos of both sexes. Since Sxl's role in supporting sex determination is not conserved outside of the *Drosophila* genus (Cline et al., 2010), it is possible that the role we have defined here resembles an ancestral one that evolved to balance fast development with proper activation of cell signaling.

## Sex-lethal is associated with truncated transcripts

If truncation of long RNAs is mediated through direct binding of Sxl, then the clusters of Sxl consensus binding sites (e.g. orange arrowheads, Figures 2I) must be transcribed for Sxl to bind and act. Using qPCR primer sets spaced along the *sog* locus (Figure S3A, blue markers) and individually staged, but not sex-selected, embryos, we found that during NC13, the *sog* intronic sequence including the Sxl binding site cluster was present, but abundance of *sog* transcript was drastically reduced in the downstream intronic sequence and the 3' coding exon (Fig. S3B,D), indicating that the truncated form, but not full-length transcript is transcribed. During NC14, the entire intron including the Sxl binding site cluster was spliced out, but the 3' coding exon was retained at high levels equivalent to the 5' exon (Figure S3C,E). At NC13 in embryos subject to maternal Sxl RNAi, more of the intron was retained, but the full transcript was still not present (Figure S3B). These results reinforce the idea that Sxl is needed for truncation of the *sog* transcript, and when Sxl is removed, truncation fails and the intron is retained.

To determine if Sxl physically associates with transcripts that exhibit truncation, we immunoprecipitated Sxl protein from a bulk collection of 2–4h embryos and performed qPCR on eluted RNA. We found that mRNAs of the genes *Sxl, msl-2*, and *tra*, which are known to be bound by Sxl for alternative splicing (Moschall et al., 2017), were enriched in the Sxl IP compared to a mock IP using a negative control antibody to Ubx, a nuclear DNA binding protein without RNA binding function as were transcripts of *sog, NetA, sca*, and *Pka-C3* (Figure 2S). Surprisingly, there was no statistical difference between the enrichment of the canonical Sxl sex-determining targets and the short transcripts investigated here. On the other hand, the genes *twi* and *sna* (short genes under 5kb) and *sog* In3B (qPCR primer 3' of the cluster of Sxl binding sites; Figure S3A), were not significantly enriched (Figure 2S), indicating little to no Sxl binding to mRNA of short genes or past the truncation point of long genes, though the intronic nature of *sog* probe In3B could lower its measured enrichment due to splicing out. These results, in combination with the presence of Sxl binding sites in the transcripts for the short forms of *sog, NetA, sca*, and *Pka-C3* genes (e.g. Data S1A–D) strongly indicate that Sxl binds to all four mRNAs found to be truncated. As short *sog* is also produced at NC14 (Figure S3F,G), the binding detected is likely a mix of Sxl protein binding to short *sog* transcript at NC13 and NC14. It is also possible that Sxl protein also associates with full-length transcripts once they are produced.

Using CRISPR/Cas9, we deleted a ~1kb region of the *sog* intron containing the Sxl binding site cluster, which we term *sog ΔSxl* (Figure 3X). When we immunoprecipitated Sxl from

embryos with this deletion and performed qPCR on associated mRNA, the association of Sxl with *sog* was greatly reduced compared to wild type, approaching the levels of negative control genes (Figure 2S). The association of Sxl with *sog* transcripts was not completely eliminated however, suggesting that while the 1kb Sxl cluster supports a significant amount of binding, other sites in the *sog* locus are likely still bound by Sxl (e.g. Fig. S1A). The association of Sxl with other mRNAs tested did not significantly change in the embryos lacking the binding site cluster in *sog*, indicating a specific interaction between Sxl and the binding sites in the *sog* intron (Figure 2S).

We also performed FISH on *sog ΔSxl* embryos using an intronic probe downstream of the deletion. In these embryos, transcriptional read-through past the truncation point was observed at NC13 with sog Intron 3 signal detection (Figure 2R), which does not occur in wild type embryos or in other controls (Figures 1C, 2Q, S3B,D,F,G,I,J), providing additional evidence that Sxl plays a key role in truncation. Furthermore, when the Sxl binding sites were mutated at the endogenous genomic locus using CRISPR-Cas9 (maintaining the spacing of the gene and FISH probes), transcriptional read-through past the truncation point into the intron was observed (Figure S2H compare with I) suggesting Sxl directly controls transcriptional termination, noting this result exhibited partial penetrance (~70% of embryos, n=5 of 7) and the extension of the transcript was observable late in NC13.

## Protein products of short transcripts are functional in signaling pathways

We investigated whether short products code for functional peptides in signaling pathways. Of particular interest, the short form of Sog contains the entire first cysteine-rich domain, which binds and sequesters TGF-β ligands Decapentaplegic (Dpp) and Screw (Scw) (Figures 3A and B; Marqués et al., 1997). The short form predicted from the 3' RACE sequence closely resembles a Sog fragment known as Supersog, both in structure and function, which was hypothesized to arise from proteolytic cleavage of full-length Sog (Yu et al., 2000). However, the 3' RACE sequence recovered for *sog* includes the use of intronic sequence as coding RNA, which would be absent from full-length Sog after splicing. Full-length Sog is cleaved by the protease Tolloid (Tld) to release ligands for signaling, but short Sog protein predicted by 3' RACE does not contain Tld cleavage sites (Peluso et al., 2011) and may bind TGF-β ligands Dpp and Scw irreversibly (Figure 3B).

To test the idea that short Sog inhibits Dpp-Scw action, we assayed the effect of ectopic expression of short Sog on the TGF-β target genes *race, hnt*, and *ush*, expressed as stripes in the dorsal ectoderm at NC14, and commonly used to assay TGF-β activity (Figures 5D,E and S4A,B; Ashe et al., 2000; Rusch and Levine, 1997). We placed the short *sog* cDNA under control of the *even-skipped* (*eve*) stripe 2 enhancer as previously done for full-length *sog* (Ashe and Levine, 1999), producing a stripe of expression along the anterior-posterior axis in addition to endogenous expression in a broad lateral domain (Figure 3C and S4F,L). In these embryos, expression of *race* is lost within the trunk and retained only in a small patch at the anterior end of the dorsal ectoderm (Figures 3H,I, compare with D,E), similar to embryos lacking functional Sog, since only the trunk expression, but not anterior domain, is Sog-dependent (Ashe and Levine, 1999; Xu et al., 2005). The expression pattern of *hnt* is also much weaker in these embryos, with the onset of expression slightly delayed, a gap in

the stripe near the posterior, and a posterior retraction from the middle of the embryo (Figure S4Q and V, compare to M and R). The expression of *ush* is weaker and slightly retracted at the anterior (Figure S4F and L, compare to A and G). These results indicate that the product of the truncated *sog* transcript, likely a short Sog peptide, acts as a dominant negative repressor of TGF-β signaling.

We also expressed *eve* stripe 2-*short sog* in a *gastrulation-defective (gd)* background, which lacks endogenous *sog* expression due to defective Toll signaling and has expanded Dpp expression throughout the embryo (Konrad et al., 1998). Concomitantly, the TGF-β pathway is activated along the entire DV axis, and *race* is ubiquitously expressed in the anterior two-thirds of the embryo (Figure 3K). In these embryos, and as shown previously (Ashe and Levine, 1999), when full-length cleavable Sog is expressed in the *eve* stripe 2 domain, robust *race* expression is observed in the anterior and mid-trunk regions, but excluded from cells expressing Sog, as it represses locally and is cleaved at a distance to activate signaling (Figure 3J). In the case of *eve* stripe 2-short *sog* in *gd*- embryos, *race* expression is limited to a band at the anterior pole of the embryo but is absent from the trunk (Figures 3 F,G compare with J,K). This result shows that short Sog does repress *race* but likely does not eliminate all signaling as expression in the head, supported by lower levels of signaling, is retained. Tld cleavage of full-length Sog is concomitant with release of ligands at a distance from the source of Sog expression and is the source of *race* expression in the trunk stripe (Ashe and Levine, 1999). In contrast, the local inhibition and lack of *race* activation at a distance in *eve* stripe 2-*short sog* embryos (Figure 3H,I,F) suggests that the predicted short Sog product cannot be cleaved by Tld to support activation of signaling and that binding of short Sog to Dpp and Scw is irreversible.

A further examination of signal transduction in the TGF-β pathway provides more evidence that short Sog sequesters the ligands and modulates signaling. The signal transducer and transcription factor Mothers against dpp (Mad) is responsible for activating transcription of TGF-β targets, and phosphorylated Mad (pMad) indicates active TGF-β signaling (Raftery and Sutherland, 1999). In wild type embryos, pMad is found in a narrow, robust, stripe along the entire dorsal ectoderm (Figures 3L,M), but in *eve* stripe 2-*short sog* embryos, pMad is diminished, ranging from decreased levels overall to mostly absent except for small patches at the anterior and posterior poles of the embryo (Figures 3N,O S4F,L). This change indicates that short Sog prevents Mad from being phosphorylated, shutting down TGF-β signaling as the retraction of the gene *race* closely matches the gap in pMad, and these changes in *race* match those observed in flies with decreased pMad (Deignan et al., 2016).

To expand our study of short Sog, we used two new and one existing mutant lines which either remove or preferentially express the short form of *sog*. Specific regions of the *sog* locus were deleted using CRISPR with the intention of disrupting or decreasing short Sog (see Methods). One deletion removed the short Sog 3' UTR sequence in the *sog* intron possibly decreasing protein levels or mRNA stability (*sog Δ New 3' UTR*), and a second deletion removed the ~1kb Sxl binding site cluster in the *sog* intron possibly leading to lack of product or longer mRNA due to defects in Sxl-mediated truncation (*sogΔSxl*; Figure 3X). In both mutants, we observed precocious and sporadic activation of *race* throughout the embryo not present in wild type embryos of the same stage (Figure 3Q and R, compare to

P). *hnt* expression in the trunk is observed earlier than in wild type (Figure S4M–O). The changes to *ush* patterns include weak early expression in the normal domain with some spots of ectopic expression that co-localize with ectopic *race* (Figure S4B,C compare with A; data not shown). This ectopic expression of *race* and *ush* and early activation of *hnt* suggest that short Sog is a dominant negative version of the protein that is important to keep cell signaling in check before cellularization, when TGF-β ligands are widely expressed throughout the embryo. Our data indicate that when levels of short Sog are altered, possibly reduced, early sequestration of the ligands fails, and TGF-β signaling is activated ectopically in the mutant.

Changes in pMad observed in the *sog ΔNew 3' UTR* and *sogΔSxl* lines help explain changes in TGF-β target genes, which are dependent on pMad for their expression. In wild type embryos, pMad is localized in a narrow band of cells in the dorsal ectoderm (Figure 3M and S4A,G), but in the two CRISPR lines, pMad is weaker in dorsal regions (Figure S4B,C). This weaker expression is likely due to a lack of Dpp concentrated at the dorsal ectoderm, spread wider throughout the embryo instead, and responsible for the precocious, ectopic *race* and *ush* expression observed in these mutant embryos (Figures 3Q,R and S4B,C). In both of the CRISPR manipulated lines, full-length *sog* is eventually transcribed later in NC14, and its activity presumably restores *race*, *ush*, and *hnt* to their usual expression domains in late NC14 (Figures 3U,V and S4H,I,S,T).

We also identified a mutant in which the *sog* locus is interrupted by a P-element insertion ~3.5kb downstream of the Sxl truncation point (Figure 3X), which causes a ~7-fold decrease in transcription of full-length Sog but allows transcription of short Sog (Figure S3H). In this genetic background, short Sog is likely intact and functional at NC13 but a deficit in long Sog occurs at NC14. In embryos with this insertion, at NC14B, when full *sog* is normally first transcribed, *race* expression is retracted to a somewhat wider anterior patch compared to wild type embryos (Figure S4D), and later in NC14, *race* is not expressed at full strength in the trunk region (Figure 3W, compare with 3T). At NC14B, *ush* is weak and slightly expanded laterally, and *hnt* expression is difficult to detect (Figure S4D,P). These results suggest that when full *sog* is available in wild type embryos to establish TGF-β signaling, the P-element line, which reduces full *sog* but allows dominant negative short *sog*, shows overall weaker expression from target genes. The phenotypes associated with the P-element at NC14B somewhat resemble those of embryos of $sog^{Y506}$ background (Ray et al., 1991), an RNA null mutant (e.g. Figure 3Z, S4E); with *race* and *ush* in $sog^{Y506}$ weaker, somewhat retracted along the AP axis, and expanded laterally. Ectopic expression of *race* at early NC14 is observed in $sog^{Y506}$ mutant embryos (Fig. 3Y) but is not present in the *sog* P-element mutant embryos (Figure 3S). These data support the view that short Sog keeps signaling off at early stages, as short *sog* is present in the P-element line, with little to no early ectopic expression, but absent in $sog^{Y506}$, where ectopic expression is observed. The similarities between the P-element and $sog^{Y506}$ diverge at late NC14, when full *sog* is available in the P-element line but not in $sog^{Y506}$, and TGF-β targets appear as normal (Figs S4 J,K,U).

The truncations we found were not limited to *sog*, and when the short peptides predicted by *NetA*, *sca*, and *Pka-C3* short transcripts (Figures 1A and 4A,D) were compared with full-

length forms, a subset of functional domains were encoded, suggesting the short forms of these genes could correspond to functional truncated proteins. By qPCR, we determined that these transcripts are truncated at NC13 with new coding sequence retained, but fully transcribed with coding sequence spliced out at NC14 (Figures 4B,C). Hydrophobicity plots of the short forms demonstrate that the new amino acids likely maintain the structure and function of the short proteins (Figure S1A–D). Previous research involving either random or targeted mutagenesis of these genes, or mammalian orthologs, has uncovered evidence of dominant negative activity in all cases at later stages of development (Hu et al., 1995; Miloudi et al., 2016; Schneiders et al., 2007).

To provide insight into the role of these other short products in embryos, we expressed short *sca* in the *eve* stripe 2 domain and looked for phenotypes in early embryos. Specifically, as Sca has been shown to form a complex with Notch and modulate its activity (Powell et al., 2001), we assayed effects on one Notch target gene *single-minded (sim)*, expressed in a thin stripe on the border of the mesoderm and neurogenic ectoderm (Figures 4E,F; rev. in Reeves and Stathopoulos, 2009). In embryos expressing *eve* stripe 2-*short sca*, *sim* is expressed early and expanded late only in the *eve* stripe 2 region, which is consistent with membrane-bound Sca protein affecting Notch locally (Figures 4G,H). In a previous study, the *sca* locus was subject to random mutagenesis, and one allele was found to have a dominant negative phenotype that affected Notch signaling (Hu et al., 1995). This allele is a truncation of the *sca* transcript just after the Rab binding domain and resembles the *short sca* truncation we recovered using 3’ RACE. It is possible that changes to *sca* shift the balance of Notch in the membrane vs. endosomes, which is mediated by Rab proteins (Hu et al., 1995).

Collectively, these data demonstrate that the long genes we observed and manipulated are truly truncated, and the full-length forms are not transcribed during NC13. Still, a recent publication has described a faster rate for RNA Pol II in *Drosophila* embryos of ~2.4 kb/min, using an analysis of heterologous engineered reporter genes of ~5kb in length (Fukaya et al., 2017). In this situation, transcription and subsequent translation of genes longer than 35kB during NC13 within 15 min would be challenging, while expression of genes less than 15kb would be achievable. Our qPCR quantification suggests long forms, if present from nascent transcription or maternal contribution, are present at ~600-fold lower levels than the short forms at NC13 (Fig. S4A,B,D Ex1:Ex5). Furthermore, we detect short transcripts present at NC14 (Fig. S4F,G) when full-length transcripts are also present suggesting that the balance of short and long forms is important for proper regulation of cell signaling.

## Global 3’ RNA-seq identifies additional truncated transcripts

To provide insight into the global or programmatic nature of transcript truncation, RNA-seq was performed on *Drosophila* embryos from NC13 and NC14 separately, targeting the 100bp at the 3’ end of transcripts (i.e. 3’ RNA-seq; Lianoglou et al., 2013). While there is little difference in 3’ transcript ends of short genes between NCs 13 and 14 (Figure 5B), long genes show large differences in 3′ transcript abundance (Figure 5A). We analyzed the dataset looking for additional short forms in NC13 examining long genes, greater than 15kb, as well as a shorter set of genes 8–15kb that are longer than average but theoretically could be transcribed within the time window available at NC13. In addition, we narrowed the

search to include only genes with mapped reads in both NC13 and NC14. Using these criteria, we manually annotated 450 genes greater than 15kb, and 354 genes 8–15kb, searching for additional short forms (Table S1). Among the 450 long genes, we found 27 putative short forms, such as the gene *grh* (Figure 5C), in addition to the four found by the original 3' RACE experiments, for a total of 31 truncated genes enriched for Gene Ontology (GO) terms Developmental Protein and Differentiation Gene (Figure 5F; Table S2; see STAR Methods). These two enriched GO functions point to a short transcript program specifically involving key developmental genes functioning in signaling and transcription in the early embryo.

In addition, many of these genes have clusters of Sxl binding sites within 1kb of their truncation points (Figure 5D). We did not find any clearly truncated genes in the 8–15kb group. This 3' RNA-seq experiment identifies global differences in truncated transcripts for both short and long genes. Moreover, our previous study using NanoString to quantify transcripts in the early embryo (Sandler and Stathopoulos, 2016b), including *sog* and *NetA*, also showed a difference in 5' vs 3' transcript abundance before NC14, confirming the results from 3' RNA-seq (Figure S3I–K).

The 3' RNA seq data also provided information on previously annotated 3' UTR usage, as a large number of genes had different 3' UTR usage between maternal and zygotic isoforms (Figure 5E). 125 of 450 long genes (i.e. >15kB) and 50 of 354 8–15kb genes had 3' peaks that were different between NC13 and NC14 (Figure 5F). All of these genes are both maternal and zygotic and using previously generated RNAseq data from staged embryos (Lott et. al., 2011), suggesting that the different 3' UTR peaks we observed corresponded with the switch from maternal to zygotic transcript in the early embryo. Moreover, the switch to zygotic 3' UTR usage, especially for long genes, occurs at NC14, when the time permissive length of the NC allows the full transcription of the zygotic form. Although likely unrelated to Sxl-mediated truncation, this observation emphasizes both the time constraints early in development and the rapid switch in transcriptional program between NC13 and NC14 during the maternal to zygotic transition.

In closing, the need to temporally regulate the quiescence and rapid initiation of signaling pathways in the embryo is critical for proper development (Ashe et al., 2000; Noordermeer et al., 1992; Queenan et al., 1997). Rapid nuclear divisions limit transcript length of key signaling pathway members (Rothe et al., 1992), but we have shown that the truncation of these long transcripts to produce short products is a mechanism used to resolve this temporal challenge to ensure the proper timing for activation and/or maintenance of signaling. In a sense, the truncation of long transcripts can be thought of as a "rescue" whereby long transcripts that would usually be degraded and lost during rapid mitotic cycles are made mature and stable by truncation, and survive to produce functional proteins. Short forms may either act as dominant negatives, like short Sog, or be constitutively active, such as short Sca. Furthermore, the shortening of transcripts and 3' UTRs has been implicated in the activation of oncogenes and the progression of cancer, in the activation of immune cells, and regulation of axon guidance (Flavell et al., 2008; Mayr and Bartel, 2009; Sandberg et al., 2008). Short transcript programs may be more widespread and important during normal development than currently appreciated.

# STAR METHODS

## Fly stocks and husbandry

All flies were reared under standard conditions at 23°C. *yw* background was used as wild type unless otherwise noted. Fly stocks used in this study are: P{His2Av-mRFP1}III.1 [Bloomington Drosophila Stock Center (BDSC)#23650], *Sxl* RNAi P{TRiP.GL00634}attP40 (BDSC #38195), *sog* P-element disruption $w^{67c23}$ P{GSV2}GS51273 (Kyoto Stock Center#207284), $gd^{7}$ (BDSC #3109), $sog^{Y506}$/FM7 ftz-lacZ (Ferguson and Anderson, 1992), $da^{1}/SM5$ (BDSC #273), $da^{k08611}/CyO$ (BDSC #12385), and *eve* Stripe 2-*sog* a gift from Hilary Ashe (Ashe and Levine, 1999). Short *sog* and short *sca* cDNA fragments were PCR amplified from cDNA reverse transcribed from embryos aged 1–3 hours using primers (see Table S3 that also introduced AscI sites on 5' and 3' ends) and subsequently cloned into the AscI site of 2s2FPE (Kosman and Small, 1997), as similarly done for full *sog* (Ashe and Levine, 1999).

Fly embryos were staged as follows for NC14:

NC14A: 5–15 min into interphase, with a 1:1 ratio of nuclear length to width, before the start of cellularization.

NC14B: 20–30 min with a nuclear elongation ratio of 2:1 and cellularization progressed <33%.

NC14C: 35–45 min with a nuclear elongation ratio of 3:1 and cellularization progressed <66%.

NC14D: 50–60 min with a nuclear elongation ratio >3:1 and cellularization progressed >66%.

For CRISPR-Cas9 mediated genome editing flies are described in the sections below.

## RNA extraction from embryos

All RNA used for RACE, NanoString, qPCR, and 3' RNA-seq was extracted from either a 2–3 hour timed collection of embryos (for RACE) or individually collected and staged embryos (for NanoString, qPCR, 3' RNA-seq) using Trizol reagent (Ambion). Timed pools of embryos were collected from apple juice plates and washed into a 1.5 ml microcentrifuge tube, excess water removed, and crushed in 1ml of Trizol Reagent (ThermoFisher). The standard Trizol protocol was followed, with the addition of a second chloroform extraction and second 70% EtOH wash. A Histone H2Av-RFP fusion was used to stage individual embryos by nuclear cycle using an epifluorescence microscope (Sandler and Stathopoulos, 2016b). Individual embryos were imaged to confirm correct nuclear cycle, snap-frozen in Trizol using liquid nitrogen, and stored at −80° C until RNA extraction.

## Generation of cDNA libraries to map transcripts

Rapid amplification of cDNA ends (RACE) libraries were created using the GeneRacer kit (ThermoFisher) for the purpose of mapping 3' ends of transcripts. Standard protocol was

followed, consisting of RNA extraction as described above, dephosphorylating mRNA using Calf Intestinal Alkaline Phosphatase (CIP), decapping mRNA using Tobacco Acid Pyrophosphatase (TAP), serial ligations of a 5' RNA oligo adapter and a 3' oligo dT adapter, and reverse transcription using Protoscript II (NEB). Extracted RNA was treated with DNase I (NEB) prior to library construction. Nested 5' and 3' RACE primers were designed to capture alternative start sites or truncations of the genes *sog, NetA, sca, Pka-C3*, and *vn*. Both 5' and 3' primers were designed to multiple exons of each gene to capture as much diversity as possible. RACE experiments were performed on RNA extracted from embryos aged 2–3 hours, which includes both NC13 and NC14. We recovered a single short isoform for each of the genes, using two separately prepared RACE libraries and sequencing eight individual RACE products per gene for both libraries. This repeated validation recovering the same short sequences for all four genes further verifies that the RACE products recovered were mature transcripts.

### NanoString assay to quantify levels of 5' and 3' ends of sog and NetA transcripts

We used NanoString technology, which directly counts mRNA transcripts using gene-specific fluorescent barcodes, without reverse transcription, fragmentation or amplification, to observe the expression of 5' and 3' ends of the genes *sog* and *NetA* (Geiss et al., 2008; Sandler and Stathopoulos, 2016b). Once extracted from individually staged embryos, total RNA was hybridized with NanoString probes at 65°C for 18 hours and then loaded onto the NanoString nCounter instrument for automated imaging and barcode counting. To normalize between embryos and allow for absolute quantification, 1$\mu$l of Affymetrix GeneChip Poly-A RNA Control was spiked into Trizol with each embryo at a dilution of 1:10000 before RNA extraction. A linear regression was made for RNA spike-in input versus counted transcripts, and all other genes were fit to the regression and quantified.

### Fluorescence in situ hybridization staining and signal quantification

Embryos aged 1–4 hours were collected and fixed using standard protocols, and Fluorescence In Situ Hybridization (FISH) was performed in order to identify transcripts in situ using labelled riboprobes following published methods (Kosman et al., 2004) but omitting Proteinase K treatment, briefly described below, To start, timed embryos were collected from apple juice plates, washed to remove yeast and debris, bleached to dechorionate, and fixed in 1:1 formaldehyde:heptane. Embryos were devitellinized and stored in MeOH at −20°C. To perform in situ hybridization, embryos were transferred to EtOH, cleared using xylenes, rehydrated and fixed in PBS, and equilibrated in hybrization solution at 55°C. Probe hybridization was done in an Eppendorf ThermoMixer C instrument at 55°C for 18 hours, gently agitating every 30 minutes. Riboprobes were synthesized using T7 RNA Polymerase and digoxigenin or biotin labeled NTP nucleotides (Roche) and a primary antibody to Histone H3 (Rabbit anti-H3, 1:10000; Abcam) was used to label histones for precise embryo staging by nuclear cycle. Embryos were sectioned along the anterior-posterior axis manually using a razor blade, and cylindrical mid-embryo sections were imaged face-on. FISH signal was quantified by normalizing signal intensity from probes to 5' and 3' ends of genes compared to signal intensity from histones in individual embryos.

### Preparation of extracts and Sxl Western blots

Extract equivalent to 16 embryos was loaded for all samples, except the 0–4hr wild-type (WT), which was loaded with 20 embryos. For unfertilized eggs and specific nuclear cycles, samples were pooled and lysed directly into 2X SDS sample buffer. Embryos from specific nuclear cycles were identified, added to the lysate pool in 2X SDS sample buffer, and snap-frozen in liquid nitrogen to prevent further development until all embryos were collected for each stage. For 0–4hr WT embryos, a large collection of embryos was taken, counted, and lysed in PBS pH7.4 with 6M urea and 1% CHAPS, incubated 10 minutes on ice, homogenized, and spun for 20 minutes to pellet debris, followed by addition of SDS lysis buffer to a 1X concentration. For the Sxl RNAi NC10–13 sample (Figure S2M), the bands in the vicinity of Sxl are somewhat warped due to a local deformation of this particular gel, but the background bands are still visible.

Extracts were separated by discontinuous denaturing 9% SDS-PAGE with AccuRuler RGB Plus/Bluestain molecular weight marker (Gold Biotechnology), and transferred to PVDF (Immobilon-P, Millipore) for Figure 2H, or BA85 Whatman Protran nitrocellulose (Supplemental Figure S2M,N) in Towbin buffer (25mM Tris, 192mM Glycine) with 5% (v/v) methanol. The membrane was rinsed extensively with dH2O, equilibrated for several minutes in TBS-T (pH 7.5 with 0.05% Tween-20), and blocked with 0.2% BSA (w/v) in TBS-T for five minutes, followed by a 10 minute TBS-T wash. The membrane was incubated overnight at 4°C with antibodies diluted in 4ml TBS-T. Mouse α-Sxl M114 (Bopp et al., 1991) was diluted 1:50, as was mouse α-BicD 1B11 (Suter and Steward, 1991). Membranes were washed 5×10 minutes, incubated with HRP-conjugated goat-α-mouse (Millipore 12–349) at 1:10,000 in TBS-T for one hour, washed as above, and rinsed extensively with TBS. The blot was developed with ProSignal Dura (Genesee Scientific) diluted 1:7 in TBS for each component, and detected with HyBlot CL film (Denville Scientific). Blots were stripped with 0.1M glycine pH 2.3 with 2% Tween-20 (v/v), and 5% SDS (w/v), washed extensively with TBS-T, reblocked as above, and reprobed.

### Immunostaining of Drosophila embryos

Concurrent immunostaining was done with in situ hybridizations using the same methods of fixation and probe hybridization as described above. Embryos were incubated in a 1:10 dilution of primary antibody supernatant (α-Sxl M114 or M18, or α-PhosphoSmad1/5) overnight at 4°C, then the antibody was washed off and embryos were incubated in a fluorescent secondary (Alexa Fluor 647 donkey α-Mouse, 1:500) for one hour at room temperature. Embryos were then washed and mounted for imaging.

### RNAi experiments using a heat-shock Gal4 approach to knock-down maternal transcripts midway through oogenesis

In most cases, the use of RNAi against or mutation of the selected RPBs causes sterility or is lethal (Johnson et al., 2010; Staller et al., 2013; Yan et al., 2014). Therefore, we employed combined heat-shock Gal4 driver with UAS-RNAi lines to generate female flies primed for RNAi (Staller et al., 2013) using an empirically devised heat-shock approach to allow the early stages of oogenesis to proceed normally and to support RNAi later in oogenesis so that maternal product in the egg would be depleted. We crossed Hsp70-GAL4 flies (BDSC

#2077) to UASRNAi line for *Sxl* (BDSC #38195). Once a stock with both components was generated, virgin females were collected and crossed back to males of the original RNAi stock. Flies were heat-shocked three days in a row at 37°C for 1.5 hours, and embryos collected on the three subsequent days. Flies from the same cross were kept without heat shock and embryos collected in parallel, as a control to confirm any phenotypes seen were due to RNAi and not non-specific effects of the constructs.

## CRISPR-Cas9 mediated genome modification

To target a deletion of the new exon or Sxl binding sites located downstream of the *sog* truncated transcript 3' end, a transgenic line was generated expressing two guide RNAs (gRNAs) targeting the region that includes the new exon or Sxl binding sites at *sog* locus. First, the unique PAM recognition sites were identified flanking this region using the flyCRISPR optimal target finder (http://tools.flycrispr.molbio.wisc.edu/targetFinder). Subsequently, these two sites were cloned into the plasmid pCFD4-U6:1_U6:3tandemgRNAs (Addgene plasmid#49411). The plasmid including these two PAM sites was injected into *$y^2cho^2v^1$; P {nos-phiC31\int.NLS}6X; attP2 (III)* (NIG-Fly #TBX-0003), resulting in phiC31-mediated site-integrated transgenesis at landing site attP2 (Chr. III) (Kondo and Ueda, 2013). Integration in the genome at this position was confirmed by PCR/sequencing.

We attempted to delete the new coding exon of short *sog*, but no PAM sequences were available, so to delete the new 3' UTR, non-homologous end joining (NHEJ) mediated by the CRISPR-Cas9 genome editing system was utilized (Kondo and Ueda, 2013). *$y^2cho^2v^1$;sp/CyO;P {nos-Cas9,$y^+$,$v^+$} 2A* (NIG-Fly #Cas-0004) virgin flies were collected and crossed with gRNA transgenic male flies. The individual progeny were screened by PCR and sequencing for the deletion (ΔNew Exon, see below). The end result is a deletion of the short Sog 3'UTR sequence that destabilizes the transcript.

>ΔNew3'UTR (black=genomic sequence, blue=introduced sequence/junction)
agtccatagcataaccattcatagcagctgccacacagaacaa

To delete the region including Sxl binding sites at the *sog* locus, homology directed repair (HDR) mediated CRISPR-Cas9 system was utilized (Gratz et al., 2014), A donor construct was generated using pHD-DsRed vector (Addgene plasmid #51434). An ~1kb 5' or 3' homology arm to the regions either upstream or downstream of the Sxl binding sites at the *sog* locus was cloned with SmaI/NheI or AscI/XhoI, respectively (creating HDR.del.sxl).

*$y^2cho^2v^1$;sp/CyO;P {nos-Cas9,$y^+$,$v^+$} 2A* (NIG-Fly #Cas-0004) virgin flies were collected and crossed with gRNA transgenic male flies. Embryos were collected and injected with 300 ng/l of the donor vector. By HDR mediated CRISPR-Cas9, an ~1.1kb region including four Sxl binding sites was replaced by a ~1.3kb fragment, which induces RFP expression in eyes (3xP3-DsRed); essentially retaining similar organization at the locus save presence of Sxl binding sites/associated sequence. The deletion of the region including Sxl binding sites was confirmed by expression of RFP in adult fly eyes and by sequencing. The RFP marker was subsequently removed by crossing the line to a Cre expressing fly line (y[1] w[67c23]

P{y[+mDint2]=Crey}1b; D[*]/TM3, Sb[1], BDSC #851). Excision of the marker was confirmed by PCR (Sxl, see below).

>ΔSxl (black=genomic sequence, blue=introduced sequence/junction, purple=loxP remnant sequence after Cre-mediated excision)

Cctattccgaatccaaatcggctagcggccgcggacatatgcacacctgcgatcgtagtgccccaactggggtaacctttgagttctc tcagttgggggcgtagataacttcgtataatgtatgctatacgaagttatagaagagcactagtaaagatctccatgcataaggcgcgc cgcgcggcttttccagcgagac

To mutate Sxl binding site at the *sog* locus, homology directed repair (HDR) mediated CRISPR-Cas9 system was utilized (Gratz et al., 2014). To mutate all four match to the Sxl consensus RNA recognition sequence of 8Us or more (see Fig. 2A; Ray et al., 2013), each corresponding nucleotide in the genomic sequence was replaced with the complementary base (i.e. A>T or C>G). 1133bp of *sog* gene intronic sequence that includes all mutated Sxl binding sites and introducing NotI and NheI sites flanking this sequence was synthesized and inserted into pUC57 (GenScript).

>mutSxl (black=genomic sequence, blue=introduced sequence/junction, purple=loxP remnant sequence after Cre-mediated excision).

cctattccgaatccaaatcggctagcggccgctggtccactacttcggataatggccacattcttgttctttttatttatttattAAAAA AAAAAAAAAAAAAtttcgttgacttttgcatttatttatttgcgtgccatgctttttcgtgtagttcgcttgctttgttttatttgat AAAAAAAAAAAAAAAAttttattttcaatctattttatatcgcccgaacggcgcctgaagttgttgctattgctgtttttgttt ctgggtttataatattatcgtggcgaatccgccgggcggtacaatgtatttcaagtatttattcgagcactttgaaggggtcccattgggg gcgcacgtgccgcattcgcaacggcttaatagaccaattaccgggataagttataaagtcgaaaactaaaaaaaaaaaaaaccgaaa gaatcaaaaattgaacaacaatcgctttctatcgtcattttcttcagctcgattgtgagcagtgtgctcggcataatttatgttcgcagtgttt tggataatttaacgcctcaattgaaaatcaaaatgggttatAAAAAAAAAAAtttcgaggcaatgtgacgaactctgtggcta ttttcactgtgacatttttcacataatcaggcgagtgctgtctgaattccagttgctgctgcatgctgcatgctgcatgttgcatgttgctgct gccttgttgccagttgctagttgccggttgctagttgccagttgccagttgctggtttactggaagttgctgtgtggcatggggcaaactg gttgccaccgaacgggaatggggttaagagacggggccggggtgatgggcgggcggaatgcggcacggcggtgcggttgtgg ggttaaggcggtcgctgcatcacatcattagtttccgttttgcggcaatttttcatttggcttatgcaaagagccgttgacccgcggacctt ccaacccgaaaacaatttcacttttccaccgctgttcatggcttttattttctcgttttttccctttacttttacttagcaatttgtttgAAAAA AAAAAACAAAAAgttttgcaccgcttccaaaaagaaaactcccaacgcaactcgtttgccataaatagttagaaggcacg gcatatgcacacctgcgatcgtagtgccccaactggggtaacctttgagttctctcagttgggggcgtagataacttcgtataatgtatg ctatacgaagttatagaagagcactagtaaagatctccatgcataaggcgcgccgcgcggcttttccagcgagac

This sequence was added to the left homology arm of the HDR donor construct used to generate the Sxl deletion (see above) following NotI/NdeI digestion, and used as donor construct in order to mutate the 4 Sxl binding sites. CRISPR-Cas9 screening to identify changed genomic sequence as well as DsRed RFP marker removal, leaving behind a loxP footprint, were conducted as described (Gratz et al., 2014). To confirm mutated sequence, genomic DNA was extracted, PCR amplified, and sequenced (mutSxl, see above).

ΔNew3'UTR, ΔSxl, and mutSxl fly stocks are viable and fertile.

### RNA IP and qPCR to assay Sxl association with transcripts

Nuclear extract preparation was based on a previously described method (Kamakaka et al., 1991). Approximately 0.4g of 2–4 hour O-R embryos were collected at 25°C and dechorionated for 3 minutes according to standard protocols in 50% bleach, washed with water, followed by a Triton-NaCl embryo wash, then rinsed with water. All following steps were performed on ice or at 4°C. Embryos were homogenized in a 2ml dounce (10 passes with pestle A, 3 passes with pestle B) in NE I (15mM HEPES pH 7.4, 10mM KCl, 5mM MgCl2, 0.2mM EDTA, and 350mM sucrose supplemented with 1x Complete protease inhibitors and PhosStop (Roche)), at a ratio of 2 ml buffer to 1g embryos. Extract was filtered through miracloth to remove debris. Nuclei were collected at 3000 × g for 10 minutes, then washed twice with NE I with gentle resuspension of nuclei, while avoiding yolk and other embryonic debris with each wash. Nuclei were then resuspended and disrupted in 150ul of NE II (50mM HEPES pH 7.4, 300mM NaCl, 0.1% Tween-20, 10% glycerol, and 0.1mM EDTA supplemented with inhibitors as in NE I) and incubated on ice for 12 minutes. The extract was spun in a microfuge at top speed for 30 minutes to remove debris.

For IP, the extract was diluted 1:1 with binding buffer (25mM HEPES pH 7.4, 10% glycerol, 1mM EDTA, 5mM KCl, and 1mg/ml BSA), using 150ul of diluted extract for each IP. Antibody-Protein G complexes were prepared by incubating 50ul of supernatants of α-Sxl (DSHB M114) or α-Ubx (DSHB Ubx/ABD-A FP6.87) in binding buffer with 30ul of Protein G beads for 1.5 hours in a total volume of 400ul, washed 2X with binding buffer, 2X with wash buffer (40mM HEPES pH 7.4, 300mM NaCl, 10% glycerol, and 0.2% NP-40), then 2X with binding buffer. Diluted nuclear extract was incubated with prepared beads with agitation for 1.5 hours, and washed 4X with wash buffer. Immunoprecipitated material was eluted with 100ul of 50mM HEPES pH 7.4, 2% Sarkosyl, and 10mM DTT for 30 minutes at 50°C. Proteinase K was added to the eluted material to a final concentration of 1mg/ml and incubated at 50°C for 30 minutes.

RNA was extracted from eluate using acid phenol:chloroform, pH 4.5 (Ambion), followed by chloroform extraction, isopropanol precipitation, and wash in 70% EtOH. RNA was treated with DNase I (NEB) and reverse transcribed using Protoscript II (NEB). qPCR was performed on cDNA using SYBR Green I Master Mix (Roche) on a StepOnePlus Real-Time PCR System (Applied Biosciences) using primers listed in Table S3. For long genes *sog, NetA, Pka-C3*, and *sca*, primers used all amplified the 5’ exons of the genes expressed as part of the short forms. Relative quantification performed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

### 3’ RNA-seq to detect global 3’ ends of genes in the embryo

RNA from pools of 50 embryos each from NCs 13 and 14 was extracted as described above. A sequencing library was created using a previously described method (Lianoglou et al., 2013) with modifications. Biotinylated oligo dT adapters with an rU residue in the dT section are were conjugated to M-280 Streptavidin Dynabeads (Invitrogen), and first and second strand cDNA synthesis were subsiquently performed with Superscript III (Invoitrogen) and DNA pol I (NEB). A single strand nick was introduced at the rU residue

using Rnase HII (NEB), and translated using E. coli DNA Pol (NEB) for eight minutes at 8°C, approximately 100 bases from the original site of the nick. DNA fragments were cleaved and blunted at the site of the translated nick with T7 Exonuclease (NEB), Mung Bean Nuclease (NEB), and Klenow DNA Pol I (NEB). Illumina TruSeq adapters were ligated onto the DNA fragments at two-fold lower concentration than the original protocol in order to reduce unincorporated adapters. The library was PCR amplified through 15 cycles, and final library was size-selected at 150–210 bp. The concentration of ligated sequencing adapters was lowered two-fold to decrease unincorporated adapters sequenced, and final library was size-selected from a 2% Ultra Pure LMP Agarose (Invitrogen), extracted from gel slices using β-Agarase I (NEB), and purified with a phenol:chloroform extraction as described above. Libraries were sequenced on an Illumina HiSeq2500 and sequenced aligned to the FlyBase (April, 2006) annotation using Tophat version 2.0.13 and Bowtie 1.1.1 as the aligner (Kim et al., 2013; Langmead et al., 2009).

RNA-seq libraries from two separate biological replicates for each nuclear cycle were prepared and sequenced independently. The first replicate was sequenced to a depth of ~25 million reads, and the second replicate was sequenced to a depth of ~150 million reads. Internally primed reads were filtered out of the aligned reads using python to build a BED file of Poly-A and Poly-T islands of at least eight bases in length, depending on sequence orientation. BEDTools was then used to intersect the BED file with the aligned reads to filter the reads within 10 bases of a Poly-A or Poly-T island (Quinlan, 2014). Internally primed reads greater than 10 bases away from a polyA stretch were not filtered out. Sequences were split based on strand orientation and seperate browser tracks created to display stranded reads, relevant to orientation of genes on positive or negative strand.

All sequence data has been uploaded to the NCBI GEO database under accession number GSE108152.

### Curation of 3'seq reads and GO analysis

All 450 genes >15kb and 354 genes 8–15kb were manually inspected, searching for signatures of short forms in the 3' RNA seq data, as seen with *sog*. Genes must have mapped reads in both NCs 13 and 14 to be included in the manual curation. 3' reads must be within 16.5kb of a transcription start site, and not within 10 bases of a poly-A stretch in the genome to be considered valid signatures of short forms. Using the DAVID Bioinformatics Gene Ontology clustering tool (Huang et. al, 2009), we found that the most enriched Gene Ontology (GO) term in 31 short forms was Developmental Protein ($p=4.8E^{-7}$), followed by Differentiation Gene ($p=2.8E^{-4}$).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank B. Williams and I. Antoshechkin for library construction and sequencing support (Millard and Muriel Jacobs Genetics and Genomics Laboratory, Caltech), C. Mayr for advice in implementing the 3' RNA-seq protocol, T. Koromila and J. McGehee for technical support, H. Ashe for sharing fly stocks, H. Araujo for providing

antibodies, and T. Cline, H. Lipshitz, and D. Rio for helpful discussions. This study was supported by NIH grant R35GM118146 to A.S. and by the Caltech Beckman Institute Functional Genomics Center (H.A.).

## REFERENCES

Ali-Murthy Z, Lott SE, Eisen MB, and Kornberg TB (2013). An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS Genet 9, e1003428. [PubMed: 23593026]

Ardehali MB, and Lis JT (2009). Tracking rates of transcription and splicing in vivo. Nat. Struct. Mol. Biol 16, 1123–1124. [PubMed: 19888309]

Artieri CG, and Fraser HB (2014). Transcript length mediates developmental timing of gene expression across Drosophila. Mol. Biol. Evol 31, 2879–2889. [PubMed: 25069653]

Ashe HL, and Levine M (1999). Local inhibition and long-range enhancement of Dpp signal transduction by Sog. Nature 398, 427–431. [PubMed: 10201373]

Ashe HL, Mannervik M, and Levine M (2000). Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo. Development 127, 3305–3312. [PubMed: 10887086]

Bell LR, Horabin JI, Schedl P, and Cline TW (1991). Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in Drosophila. Cell 65, 229–239. [PubMed: 2015624]

Bopp D, Bell LR, Cline TW, and Schedl P (1991). Developmental distribution of female-specific Sex-lethal proteins in Drosophila melanogaster. Genes Dev 5, 403–415. [PubMed: 1900493]

Cline TW (1993). The Drosophila sex determination signal: how do flies count to two? Trends Genet 9, 385–390. [PubMed: 8310535]

Cline TW, Dorsett M, Sun S, Harrison MM, Dines J, Sefton L, and Megna L (2010). Evolution of the Drosophila feminizing switch gene Sex-lethal. Genetics 186, 1321–1336. [PubMed: 20837995]

Cronmiller C, and Cline TW (1987). The Drosophila sex determination gene daughterless has different functions in the germ line versus the soma. Cell 48, 479–487. [PubMed: 3802198]

Deignan L, Pinheiro MT, Sutcliffe C, Saunders A, Wilcockson SG, Zeef LAH, Donaldson IJ, and Ashe HL (2016). Regulation of the BMP Signaling-Responsive Transcriptional Network in the Drosophila Embryo. PLoS Genet 12, e1006164. [PubMed: 27379389]

Ferguson EL, Anderson KV, Localized enhancement and repression of the activity of the TGF- family member, decapentaplegic, is necessary for dorsal-ventral pattern formation in the Drosophila embryo. Development 114, 1992, 583–597. [PubMed: 1618130]

Flavell SW, Kim T-K, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, and Greenberg ME (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60, 1022–1038. [PubMed: 19109909]

Foe VE, and Alberts BM (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J. Cell Sci 61, 31–70. [PubMed: 6411748]

Fukaya T, Lim B, and Levine M (2017). Rapid Rates of Pol II Elongation in the Drosophila Embryo. Curr. Biol 27, 1387–1391. [PubMed: 28457866]

Garcia HG, Tikhonov M, Lin A, and Gregor T (2013). Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr. Biol 23, 2140–2145. [PubMed: 24139738]

Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat. Biotechnol 26, 317–325. [PubMed: 18278033]

Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, and O'Connor-Giles KM (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 196, 961–971. [PubMed: 24478335]

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature 471, 473–479. [PubMed: 21179090]

Hoskins RA, Landolin JM, Brown JB, Sandler JE, Takahashi H, Lassmann T, Yu C, Booth BW, Zhang D, Wan KH, et al. (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res 21, 182–192. [PubMed: 21177961]

Hu X, Lee EC, and Baker NE (1995). Molecular analysis of scabrous mutant alleles from Drosophila melanogaster indicates a secreted protein with two functional domains. Genetics 141, 607–617. [PubMed: 8647397]

Johnson ML, Nagengast AA, and Salz HK (2010). PPS, a large multidomain protein, functions with sex-lethal to regulate alternative splicing in Drosophila. PLoS Genet 6, e1000872. [PubMed: 20221253]

Käll L, Krogh A, and Sonnhammer ELL (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol 338, 1027–1036. [PubMed: 15111065]

Kamakaka RT, Tyree CM, and Kadonaga JT (1991). Accurate and efficient RNA polymerase II transcription with a soluble nuclear fraction derived from Drosophila embryos. Proc. Natl. Acad. Sci. U. S. A 88, 1024–1028. [PubMed: 1992453]

Keyes LN, Cline TW, and Schedl P (1992). The primary sex determination signal of Drosophila acts at the level of transcription. Cell 68, 933–943. [PubMed: 1547493]

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, and Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36. [PubMed: 23618408]

Kondo S, and Ueda R (2013). Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics 195, 715–721. [PubMed: 24002648]

Konrad KD, Goralski TJ, Mahowald AP, and Marsh JL (1998). The gastrulation defective gene of Drosophila melanogaster is a member of the serine protease superfamily. Proc. Natl. Acad. Sci. U. S. A 95, 6819–6824. [PubMed: 9618496]

Kosman D, and Small S (1997). Concentration-dependent patterning by an ectopic expression domain of the Drosophila gap gene knirps. Development 124, 1343–1354. [PubMed: 9118805]

Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, and Bier E (2004). Multiplex detection of RNA expression in Drosophila embryos. Science 305, 846. [PubMed: 15297669]

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25. [PubMed: 19261174]

Lianoglou S, Garg V, Yang JL, Leslie CS, and Mayr C (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev 27, 2380–2396. [PubMed: 24145798]

Livak KJ, and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. [PubMed: 11846609]

Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, and Eisen MB (2011). Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol 9, e1000590. [PubMed: 21346796]

Marqués G, Musacchio M, Shimell MJ, Wünnenberg-Stapleton K, Cho KW, and O'Connor MB (1997). Production of a DPP activity gradient in the early Drosophila embryo through the opposing actions of the SOG and TLD proteins. Cell 91, 417–426. [PubMed: 9363950]

Mayr C, and Bartel DP (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138, 673–684. [PubMed: 19703394]

McLeay RC, and Bailey TL (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 11, 165. [PubMed: 20356413]

Miloudi K, Binet F, Wilson A, Cerani A, Oubaha M, Menard C, Henriques S, Mawambo G, Dejda A, Nguyen PT, et al. (2016). Truncated netrin-1 contributes to pathological vascular permeability in diabetic retinopathy. J. Clin. Invest 126, 3006–3022. [PubMed: 27400127]

Moschall R, Gaik M, and Medenbach J (2017). Promiscuity in post-transcriptional control of gene expression: Drosophila sex-lethal and its regulatory partnerships. FEBS Lett 591, 1471–1488. [PubMed: 28391641]

Noordermeer J, Johnston P, Rijsewijk F, Nusse R, and Lawrence PA (1992). The consequences of ubiquitous expression of the wingless gene in the Drosophila embryo. Development 116, 711–719. [PubMed: 1289061]

O'Farrell PH (1992). Big genes and little genes and deadlines for transcription. Nature 359, 366–367. [PubMed: 1406945]

Paz I, Kosti I, Ares M Jr, Cline M, and Mandel-Gutfreund Y (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res 42, W361–W367. [PubMed: 24829458]

Peluso CE, Umulis D, Kim Y-J, O'Connor MB, and Serpe M (2011). Shaping BMP morphogen gradients through enzyme-substrate interactions. Dev. Cell 21, 375–383. [PubMed: 21839924]

Powell PA, Wesley C, Spencer S, and Cagan RL (2001). Scabrous complexes with Notch to mediate boundary formation. Nature 409, 626–630. [PubMed: 11214322]

Pritchard DK, and Schubiger G (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev 10, 1131–1142. [PubMed: 8654928]

Queenan AM, Ghabrial A, and Schüpbach T (1997). Ectopic activation of torpedo/Egfr, a Drosophila receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. Development 124, 3871–3880. [PubMed: 9367443]

Quinlan AR (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics 47, 11.12.1–34.

Raftery LA, and Sutherland DJ (1999). TGF-beta family signal transduction in Drosophila development: from Mad to Smads. Dev. Biol 210, 251–268. [PubMed: 10357889]

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature 499, 172–177. [PubMed: 23846655]

Ray RP, Arora K, Nüsslein-Volhard C, and Gelbart WM (1991). The control of cell fate along the dorsal-ventral axis of the Drosophila embryo. Development 113, 35–54. [PubMed: 1765005]

Reeves GT, and Stathopoulos A (2009). Graded dorsal and differential gene regulation in the Drosophila embryo. Cold Spring Harb. Perspect. Biol 1, a000836. [PubMed: 20066095]

Reeves GT, Trisnadi N, Truong TV, Nahmad M, Katz S, and Stathopoulos A (2012). Dorsal-ventral gene expression in the Drosophila embryo reflects the dynamics and precision of the dorsal nuclear gradient. Dev. Cell 22, 544–557. [PubMed: 22342544]

Rothe M, Pehl M, Taubert H, and Jäckle H (1992). Loss of gene function through rapid mitotic cycles in the Drosophila embryo. Nature 359, 156–159. [PubMed: 1522901]

Rusch J, and Levine M (1997). Regulation of a dpp target gene in the Drosophila embryo. Development 124, 303–311. [PubMed: 9053307]

Salz HK, and Erickson JW (2010). Sex determination in Drosophila: The view from the top. Fly 4, 60–70. [PubMed: 20160499]

Sandberg R, Neilson JR, Sarma A, Sharp PA, and Burge CB (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 320, 1643–1647. [PubMed: 18566288]

Sandler JE, and Stathopoulos A (2016a). Stepwise Progression of Embryonic Patterning. Trends Genet 32, 432–443. [PubMed: 27230753]

Sandler JE, and Stathopoulos A (2016b). Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. Genetics 202, 1575–1584. [PubMed: 26896327]

Schneiders FI, Maertens B, Böse K, Li Y, Brunken WJ, Paulsson M, Smyth N, and Koch M (2007). Binding of Netrin-4 to Laminin Short Arms Regulates Basement Membrane Assembly. J. Biol. Chem 282, 23750–23758. [PubMed: 17588941]

Shermoen AW, and O'Farrell PH (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell 67, 303–310. [PubMed: 1680567]

Shimmi O, and O'Connor MB (2003). Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the Drosophila embryo. Development 130, 4673–4682. [PubMed: 12925593]

Staller MV, Yan D, Randklev S, Bragdon MD, Wunderlich ZB, Tao R, Perkins LA, Depace AH, and Perrimon N (2013). Depleting gene activities in early Drosophila embryos with the “maternal-Gal4-shRNA” system. Genetics 193, 51–61. [PubMed: 23105012]

Stathopoulos A, and Levine M (2004). Whole-genome analysis of Drosophila gastrulation. Curr. Opin. Genet. Dev 14, 477–484. [PubMed: 15380237]

Suter B, and Steward R (1991). Requirement for phosphorylation and localization of the Bicaudal-D protein in Drosophila oocyte differentiation. Cell 67, 917–926. [PubMed: 1959135]

Tadros W, and Lipshitz HD (2009). The maternal-to-zygotic transition: a play in two acts. Development 136, 3033–3042. [PubMed: 19700615]

Wharton SJ, Basu SP, and Ashe HL (2004). Smad affinity can direct distinct readouts of the embryonic extracellular Dpp gradient in Drosophila. Curr. Biol 14, 1550–1558. [PubMed: 15341741]

Xu M, Kirov N, and Rushlow C (2005). Peak levels of BMP in the Drosophila embryo control target genes by a feed-forward mechanism. Development 132, 1637–1647. [PubMed: 15728670]

Yan D, Neumüller RA, Buckner M, Ayers K, Li H, Hu Y, Yang-Zhou D, Pan L, Wang X, Kelley C, et al. (2014). A Regulatory Network of Drosophila Germline Stem Cell Self-Renewal. Dev. Cell 28, 459–473. [PubMed: 24576427]

Yu K, Srinivasan S, Shimmi O, Biehs B, Rashka KE, Kimelman D, O’Connor MB, and Bier E (2000). Processing of the Drosophila Sog protein creates a novel BMP inhibitory activity. Development 127, 2143–2154. [PubMed: 10769238]

## Highlights

Sandler et al. identify a developmental program where long genes are expressed during short nuclear cycles in the early *Drosophila* embryo as truncated short transcripts to temporally control signaling and developmental progression. The RNA binding protein Sex-lethal directly promotes short transcript generation in embryos of both sexes.

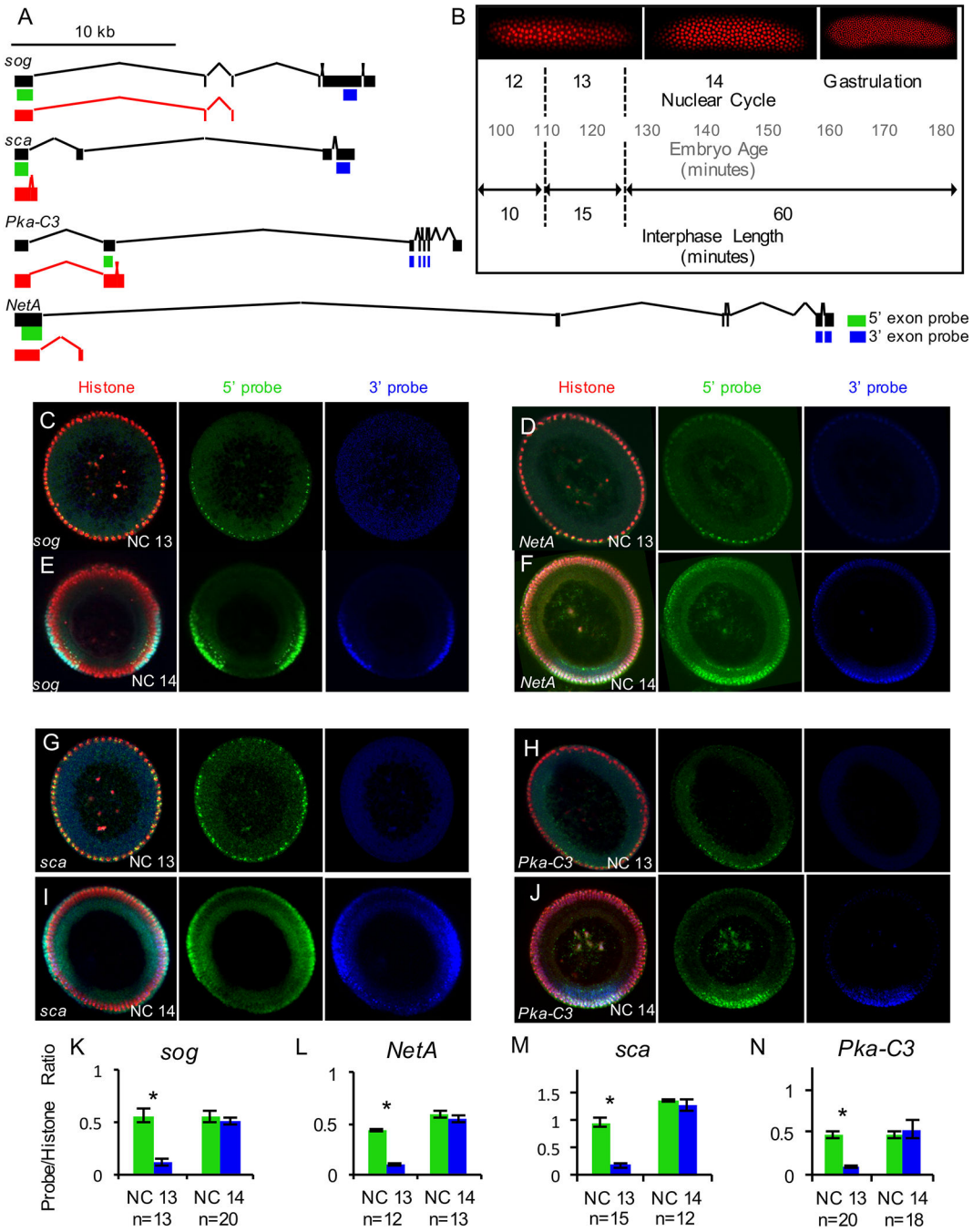



**Figure 1. Long genes are transcribed as short forms in NC13.**

(A) Full-length transcripts (black) and mapped 3' RACE identified shorter transcripts (red) for genes investigated. Locations of 5' and 3' FISH riboprobes shown in green and blue, respectively.

(B) A timeline of the syncytial blastoderm development, showing age of embryo, nuclear cycle, and interphase length for the last three syncytial nuclear cycles. Embryo images illustrate rapid division of nuclei using a Histone H2Av-RFP fusion line.

(C-J) FISH using 5’ and 3’ riboprobes for the genes *sog* (C and E), *NetA* (D and F), *sca* (G and I), and *Pka-C3* (H and J) and anti-Histone H3 antibody shown for embryos of stage NC13 and NC14. Images depict manually chopped embryo cross-sections, as described in the methods, stained to show histones (red), 5’ probes (green), and 3’ probes (blue). (K-N) Normalization of 5’ and 3’ FISH riboprobe stainings of genes *sog* (K), *sca* (L), *Pka-C3* (M), and *NetA* (N) to immunostained Histone H3 to compare signal intensity, with number of embryos analyzed for each gene and nuclear cycle below the charts. Differences are present for all genes in NC13 Data are presented as means ± SEM. Asterisks specify $p<0.0001$, two-tailed Student’s t-test. (See also Data S1)

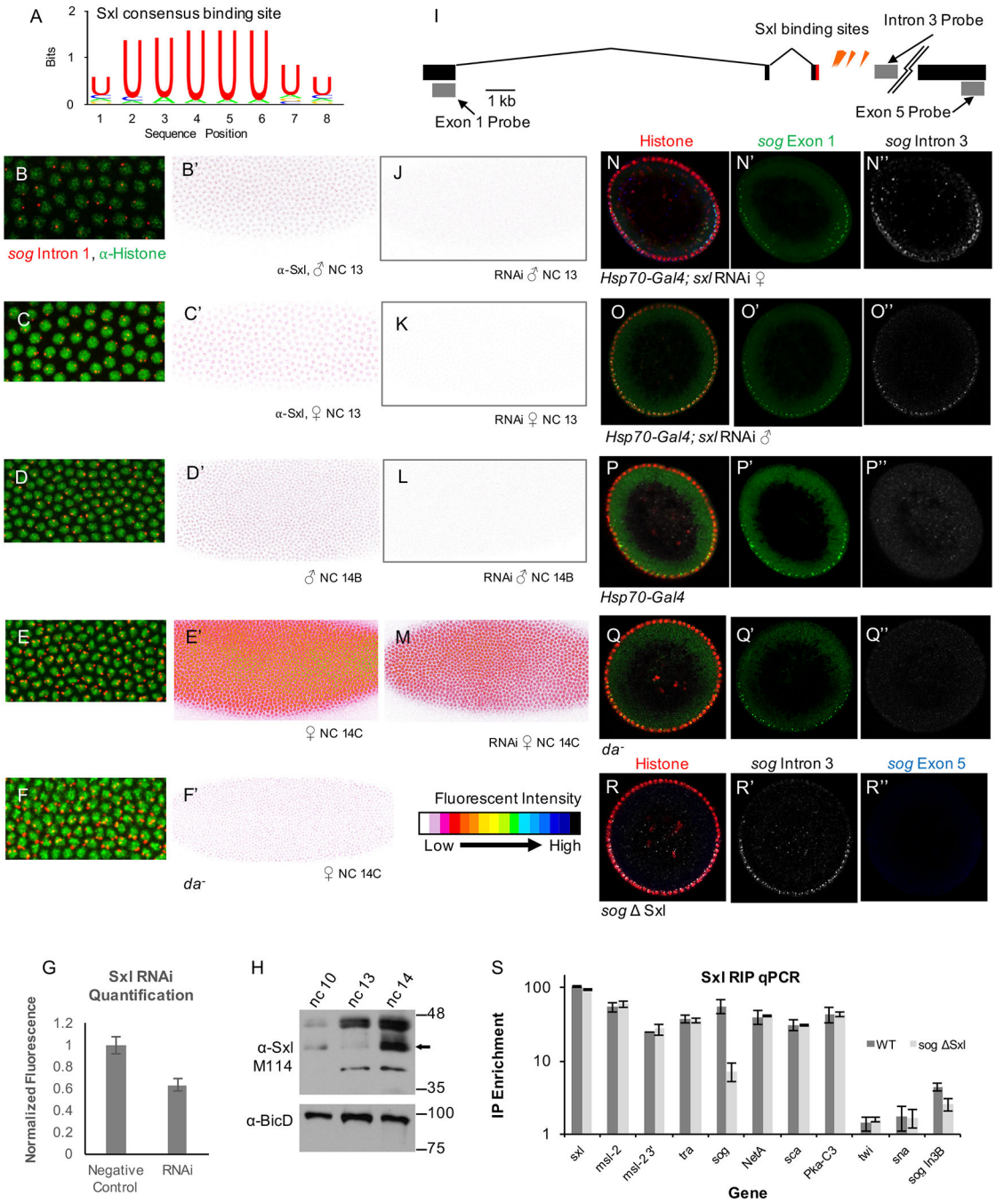



**Figure 2. Early embryonic Sxl is present in early embryos and its depletion leads to defective truncation of long transcripts.**

(A) A position weight matrix for the Sxl consensus binding site on mRNA (Ray et al., 2013).

(B-F) Immunostaining coupled with in situ hybridization to detect both protein and nascent transcripts in wild type (B-E) or *da*- (F) embryos. Embryos were stained with anti-histone H3 (green) and FISH using *sog* Intron 1 riboprobe (red) in order to determine nuclear cycle and sex based on nuclear density and number of nuclear dots (I-M), as well as monoclonal anti-Sxl antibody M114 to determine levels of this protein (I'-M', see also fluorescent heat

map key lower right). Panels in I-M are magnified views of same embryos shown in I'-M'.
(B-E) FISH in NC13 embryos of *Sxl* RNAi (B,C), negative control (D), or *da* mutant ($da^{1}/da^{k08611}$) (E) backgrounds showing signals for *sog* Exon 1 riboprobe in green, *sog* Intron 3 riboprobe in white. Number of nascent nuclear spots for *sog* genee (on X) also show embryo in C is female and that in D is male. Over 100 individual Sxl immunostained embryos were imaged for these experiments.
(G) Quantification of Sxl RNAi in immunostained embryos in which heat-shock mediated Gal4 expression was used to induce maternal Sxl RNAi or to drive LacZ expression, which served as negative control. Sxl RNAi (n=5) or LacZ positive (n=5) embryo images were analyzed using Fiji software. Embryos were outlined and average fluorescence intensity was calculated. Error bars represent SEM, Student's t-test p=0.0005.
(H) Western blot using M114 monoclonal Sxl antibody to probe extracts of 12 embryos each of NC10, NC13, and NC14. Blot was stripped and reprobed with anti-BicD as a loading control.
(I) A diagram of the *sog* locus with in situ probe locations (grey boxes), short form coding sequence (red box), and Sxl binding site locations (orange arrowheads).
(J–M) Immunostaining for Sxl in *sxl* RNAi embryos, showing a large reduction of maternal Sxl in both male and female NC13 embryos (J and K) and male NC14 (L) embryos, and moderate reduction of Sxl in female NC14 embryos (M), as zygotic expression counteracts RNAi effect. Number of embryos imaged as follows: yw NC13: 47, yw NC14: 110, RNAi NC13: 14, RNAi NC14: 25, da- NC13: 7, da- NC14: 14.
(R) FISH detection of *sog* Intron 3 (white) and Intron 5 (blue) with Histone H3 (red) in a *sog ΔSxl* mutant embryo of stage late NC13.
(S) Sxl RNA Immunoprecipitation (RIP)-qPCR (see Methods) on RNA eluted from wild type embryos (dark grey) and *sogΔSxl* embryos (light grey) using either antibodies to Sxl or Ubx (unrelated factor, mock sample) to immunoprecipitate. Transcripts assayed by qPCR include Sxl sex target genes (*Sxl*, *msl-2*, and *tra*), four truncated genes identified in this study (*sog, NetA, sca, Pka-C3*), short genes (*twi*, *sna*) as well as *sog* transcript past the short-isoform truncation point (assayed by *sog* In3B probe downstream of Sxl binding sites). Data are presented as means ± SEM. Asterisk indicates p<0.0001, two-tailed Student's t-test. (See also Data S1 and Figures S2 and S3)

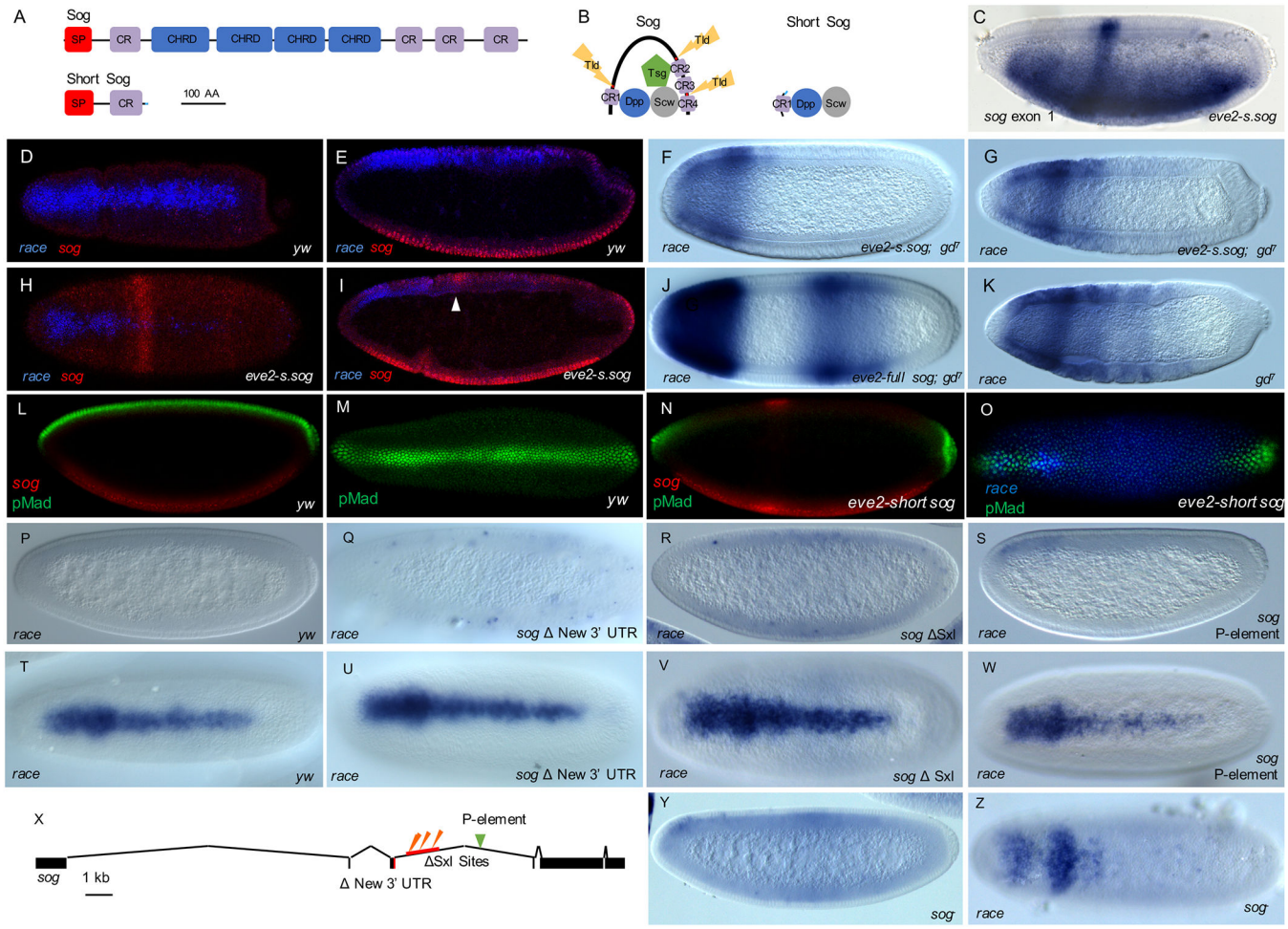



**Figure 3. Short Sog protein acts in a dominant negative manner to regulate onset of TGF-β signaling in the embryo.**

(A) Full Sog protein and predicted Short Sog protein product based on 3' RACE sequence, with functional domains shown. SP=signal peptide, CR=Cystine repeat domain, CHRD=Chordin domain.

(B) Full Sog and Short Sog bound to TGF-β ligands Dpp and Scw, along with binding partner Twisted gastrulation (Tsg) and protease Tolloid (Tld) (Shimmi and O'Connor, 2003). Short Sog lacks Tld cleavage sites (red in Sog), and instead contains amino acids from intron-derived sequence (blue; see also Figure S1A).

(C) In situ hybridization in *eve* stripe 2-*short sog* embryo, mid NC14, using riboprobe to first exon of *sog*, which recognizes both short and long forms of *sog*, and demonstrates ectopic expression of short *sog* transcript in the *eve* stripe 2 domain.

(D,E,H,I) FISH colocalization using riboprobes to the TGF-β target gene *race* (blue) as well as *sog* (red) at NC14D in wild type embryos (D,E) or *eve2-short sog* background (H,I). Dorsal (D,H; surface plane) and lateral (E,I; sagittal plane) views show single stripe of *race* gene expression present dorsally in wildtype embryos (D,E), while in comparison *race* expression in *eve2-short sog* background is diminished and excluded from the trunk (H,I). *sog* expression is only apparent in lateral view of wildtype embryos (E), but *eve2-short sog* construct supports additional expression in a stripe at the anterior (H and I, see arrowhead).

(F,G,J,K) Expression of *race* in $gd^7$ NC14C (F,J) and or gastrulating (G,K) embryos from *eve* stripe 2-*short sog* background (F,G), *eve* stripe 2- *sog* background (J), or native background, not containing any transgene (K).

(L-O) FISH detection using riboprobes to *sog* (red) or *race* (blue) coupled with anti-pMad immunostaining (green) in wild type (L, M) or *eve* stripe 2-*short sog* (N,O) embryos. Lateral (L,N) and dorsal (M, O) views are shown. Images are confocal single scans of the embryos' surface at NC14C-D.

(P-W,Y,Z) *race* expression in *yw* (P,T), *sog ΔNew 3' UTR* embryos (Q,U), *sog* (R,V), *sog* P-element (S,W), and $sog^{Y506}$ mutant (Y,Z) embryos. Shown are lateral views of NC14 early stage (P,Q,R,S,Y) or dorsal views of NC14 late stage (T,U,V,W,Z) embryos. Precocious expression of *race* (Q,R,Y) correlates with loss of short *sog* activity.

(X) A diagram of the *sog* locus with sites of genomic manipulation noted. *ΔNew 3' UTR* is the CRISPR-Cas9 deletion of the *short sog* coding sequence derived from the intron. ΔSxl sites is the CRISPR-Cas9 deletion of ~1kb of genomic DNA containing the Sxl binding sites. Green arrowhead indicates the insertion location of *sog* P-element: P{GSV2}GS51273.

(See also Figure S4)

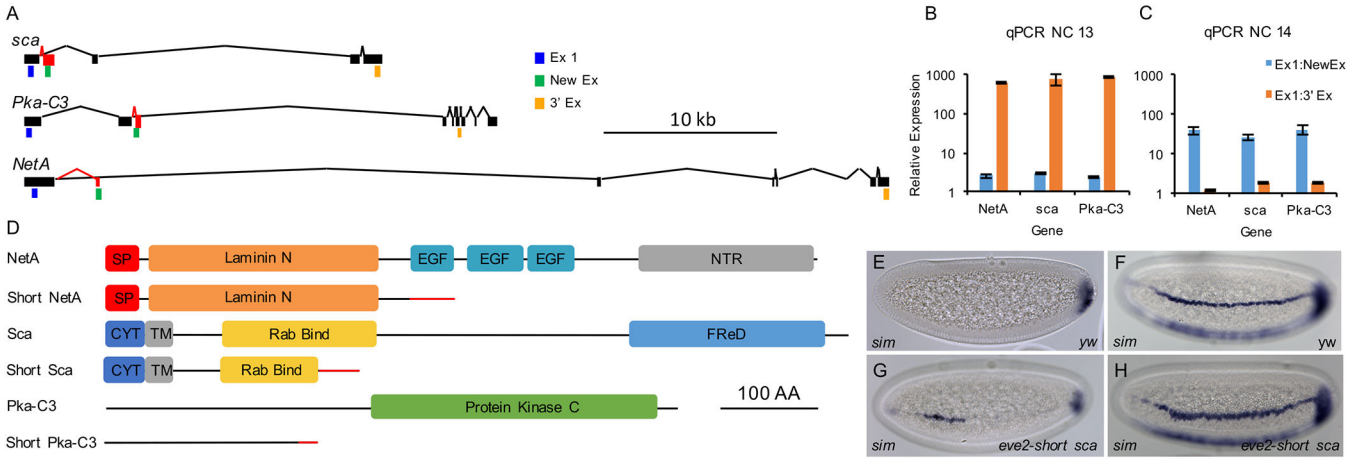


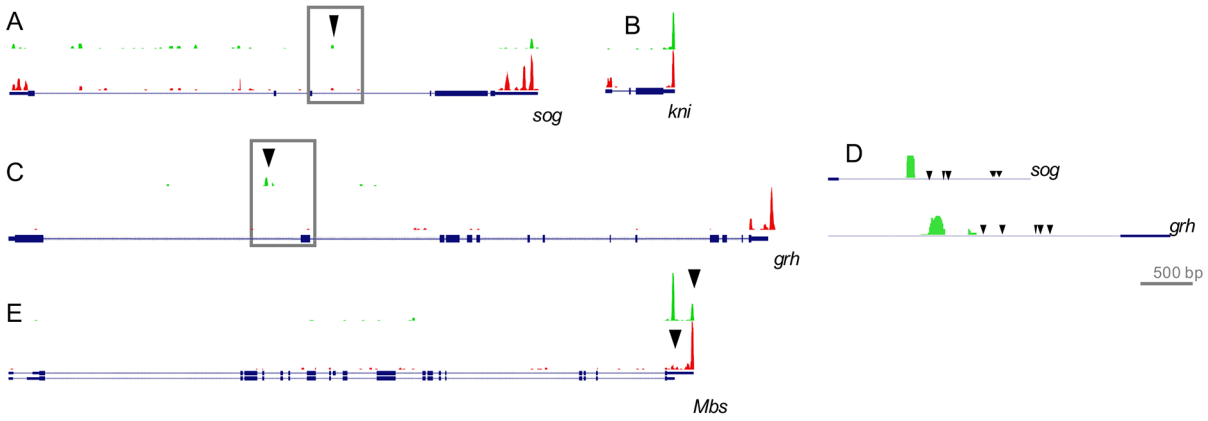

**Figure 4. Differential expression and ectopic expression of short forms for additional genes.**

(A) Diagrams of the full-length forms (black) and short forms (red) of the genes *sca, Pka-C3*, and *NetA*. qPCR primer locations are also indicated in blue for common 5' exon (Ex1), green for short form specific exon (New Ex), and orange for long form specific 3' exon (3' Ex).(B,C) qPCR of eight individually staged NC13 (B) or NC14 (C) embryos, comparing expression of 5' exons (Ex1) to both short form specific exons (NewEx) and 3' exons (3' Ex) for *NetA*, *sca*, and *Pka-C3*. Error bars represent SEM.

(D) Diagrams of functional domains for full length and short forms of proteins for Sca, NetA, and Pka-C3. SP=signal peptide; NTR=netrin domain; CYT=cytoplasmic domain; TM=transmembrane domain; Rab Bind=Rab protein binding domain; FReD=fibrinogen C-terminal domain; and other domains as marked.

(E-H) In situ hybridization using a riboprobe to detect *sim* expression in either wildtype embryos (E,F) or embryos expressing *eve2-short sca* at early (E,G) and late (F,H) NC14.

**(**See also Figures S2)




F

| | Genes expressed | Genes truncated | Percent truncated |
|---|---|---|---|
| Long (>15kb) | 450 | 31 | 7.1 |
| Short (8-15kb) | 354 | 0 | 0 |

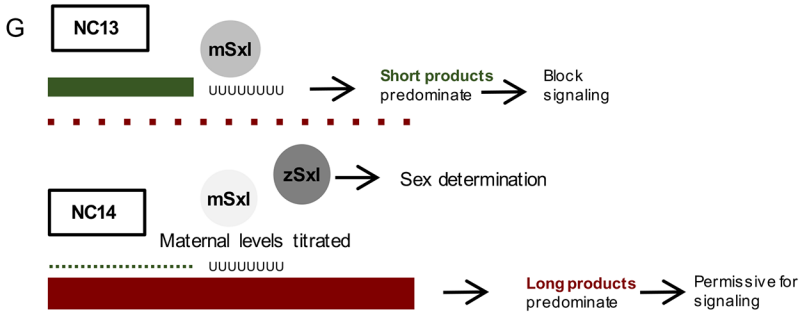



**Figure 5. Global 3' RNA-seq identifies genes with different transcripts lengths in NC13 versus NC14.**

(A-C) 3' RNA-seq browser tracks for the genes *sog* (A), *kni* (B), and *grh* (C) with sequencing reads for NC13 samples displayed in green and those for NC14 samples in red. Black arrows in *sog* and *grh* indicate locations of short form reads.

(D) Zoomed region of short form truncation points for *sog* and *grh*, with black arrowheads indicating Sxl binding sites.

(E) 3' RNA-seq browser track for the gene *Mbs*, showing maternal and zygotic isoforms and the switch in 3' UTR from NC13 to NC14.

(F) Analysis of the 3' RNA-seq displayed in a table displaying number of medium genes (8–15kb) and long genes (>15kb) expressed in both NC13 and NC14, in comparison with number of genes for each class truncated to a shorter form in NC13 compared to length in NC14 ("Genes truncated NC13 vs. NC14"), followed by calculation of percent of total number of genes of Short or Long class designation exhibiting such a truncation ("Percent truncated").

(G) Summary of mechanism by which maternal Sxl supports the generation of short transcripts and their requirement to temporally block signaling.

(See also Figure S3 and Tables S1,S2)